Vocabularies and Common Data Elements Workspace
Meeting Record
100 - 300 PM EST
November 04, 2004

Teleconference Information:
800.369.6099
Passcode: 14469
Moderator: Christine Richardson

Net-Centra Information:
http://ncicb.centra.com
Meeting ID: QKK481718

Planned Key Decisions and/or Outcomes: *(copy & paste from agenda)*	A discussion of the caBIG Business Process will occur in order to ensure that the WS participants are conversant with this topic. Additionally, we will be reviewing decisions and action items from the joint Architecture/VCDE Face-to-Face meeting.
Executive Summary:	The order of the agenda was changed and the meeting began with liaison reports from the other WS. Next, VCDE participants who attended the joint Architecture/VCDE face-to-face meeting provided a high level overview of decisions made at the face-to-face meeting. This recap evolved into an in depth discussion on vocabulary development and deployment. The group reached consensus on: • The need to address vocabulary development and deployment as two separate issues • The need to have multiple terminology servers located around the grid Due to time constraints the caBIG Business Process discussion was tabled.

Action Items/Next Steps:

Name Responsible	Action Item	Date Due
Harold Solbrig	Contact caGRID team to discuss LexGrid servers in the context of caGRID	11/12
VCDE WS	Identify/define use cases of application usage of vocabulary to drive vocabulary deployment issues	
VCDE WS	Identify/define use cases from data consumer perspective and enumerate some of the uses of data to drive vocabulary development issues	
VCDE WS	Identify a minimal set of requirements for semantic rigor in order to review and approve local vocabularies for caBIG use	
Christine Richardson/Mike Keller	Contact individual Cancer Centers to discuss caBIG Business Process	11/12

Vocabularies and Common Data Elements Workspace
Meeting Record
100 - 300 PM EST
November 04, 2004

Attendees:	Name	Affiliation
	Quan Chen	Albert Einstein
	Xin Zheng	Albert Einstein
	David Aronow	Ardais
	Hemant Shah	City of Hope
	Chris Abajian	Fred Hutchinson
	Michelle Smerek	Fred Hutchinson
	Debbie Krupke	Jackson Lab
	Jim Kadin	Jackson Lab
	Martin Ringwald	Jackson Lab
	Lyubov Remennick	Kevric
	Harold Solbrig	Mayo
	Tom Johnson	Mayo
	Veena Rajaraman	OHSU
	Lara Fournier	OHSU
	Peter Covitz	NCICB
	George Komatsoulis	NCICB
	Denise Warzel	NCICB
	Christo Andyondis	NCICB
	Brenda Duggin	NCICB
	Frank Hartel	NCICB
	Margaret Haber	NCI/OC
	Larry Wright	NCI/OC
	Cecil Lynch	UC-Davis
	Lynne Wilkens	University of Hawaii
	Leo Cheung	University of Hawaii
	Rebecca Crowley	UPMC
	Christine Richardson	VCDE WS-caBIG Project Team
	Mike Keller	VCDE WS-caBIG Project Team
	Arumani Manisundaram	Arch WS-caBIG Project Team

Vocabularies and Common Data Elements Workspace
Meeting Record
100 - 300 PM EST
November 04, 2004

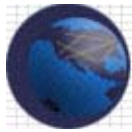| **Detailed Meeting Notes:**	
(copy & paste from agenda and place your meeting notes under each agenda item)	**Liaison Reports**
	TBPT WS (Rebecca Crowley, Larry Wright)
	• Most recent TBPT WS teleconference provided an overview of caTISSUE Lite Project
	o caTISSUE-6-8 month time frame
	▪ caTISSUE Lite is the core with additional key modules (such as an annotation module) composing caTISSUE
	o John Gilbertson will be responsible for the VCDE SIG in TBPT
	ICR WS (Quan Chen)
	• Data exchange format decisions made independent of data standards
	• Relationship between content and format standard
	o Not fully answered
	o Requires VCDE/Architecture interaction
	▪ Review mechanism would kick-in
	o Need to use as test cases
	o If agreement is reached for an exchange format, then the VCDE WS can assist with content standard
	▪ Can also insert additional requirements for ontologies
	CTMS WS (Tom Johnson)
	• Current topics under discussion:
	o Use of LOINC Codes
	o Role of caDSR, EVS
	o HL7v2 messaging moving to HL7v3 with adapters
	Training WG
	• Discussed templates and nature of templates
	• Software Documentation
	• Exploratory call held to define what is required

- o   Re-launch of Training

- Discussion of caDSR Training

    - o   1.5 hour sessions

    - o   Not recording the first round of training

    - o   If someone wants to jump into training late, they can, need to review previous slides posted on the training

    - o   Needs more of an in-depth presentation of the ISO/IEC11179 model

Recap of Joint Architecture/VCDE Face-to-Face Meeting

Various VCDE members who attended the Joint Architecture/VCDE Face-to-face meeting provided a recap of the topics covered at the meeting.

- **Metadata Representation**

    - o   XML/XML-S vs. RDF/RDF-S

        - ▪   Investigating if XML will be sufficient to represent metadata at run-time.

        - ▪   Whichever format is chosen will serve as the bridge between the object models and OWL (the semantic tier).

        - ▪   At the W3C meeting, there was a very strong RDF focus. Agreed to revisit where RDF is going.

- **Data Provenance**

    - o   Minimal provenance model about services and data

    - o   W3C follow-up

        - ▪   Metadata/RDF provenance

        - ▪   Annotia

            - •   Provenance for RDF projects underway

            - •   This is part of MyGRID in Manchester

- **Identifier Management**

    - o   W3C Follow-up

        - ▪   LSID not discussed much

        - ▪   LSID open to developing spec, open to feedback

        - ▪   Potentially use LSID as the core and extend as

- necessary for caBIG
    - LSID and HIPAA controlled information?
        - Opaque tokens for de-identified data?
        - Can be within system, but anything that goes out cannot have any individual-identifying information
    - LSID creators are essentially under the impression that it is done
  - Kick-off meeting of Identifier SIG week of Nov. 15th
  - ID SIG will take recommendation to caGRID team to determine how to implement
  - If major changes are needed for LSID to work, then we will need another technology

- **Common Query Language**
  - The caGRID team is tasked with determining common query language to use and report back
    - Suggested to look at SPARKLE (spoke about at W3C)
        - This is a successor to RQL or RDQL

- **Data and Vocabulary Standards**
  - VCDE final approval of data standards and terminologies
  - Sounding and approval board for standards and making sure silos do not arise in caBIG or in other NCICB Contexts
  - HL7 Datatypes on caDSR

- **Vocabulary Requirements**
  - There was little progress made on this topic, however it was a very good airing out of issues
  - A strategy is needed to move forward on this topic
  - There were actually two issues being discussed:
    - Vocabulary Deployment
    - Vocabulary Development
  - It was agreed that these two issues must be discussed separately in order to make progress
  - Vocabulary Deployment

- Need to address the need for standard, unified services anywhere on the grid

- Where and how do you set up vocabulary services on the grid?

- How do you want to deploy at run-time?

- What is the strategy and how does it tie in?

- Mayo and EVS terminology servers need to be adapted to conform to grid specs

- What will the landscape of servers be and what is the strategy to solve this problem?

- How does the user access servers?

- Consensus was reached that multiple terminology servers around the grid would be needed

- Globus terminology service has an adapter that can link LexGrid Service and servers

  - Harold will contact Arumani and the caGRID team to discuss LexGrid servers in the context of caGRID

- Access Model

  - Cached locally?

  - Support a variety of different approaches

  - Caching of data changing weekly vs. minute to minute?

o Vocabulary Development

- A major issue is how do deal with local vocabularies, that have never been published and what to do when ontologies, vocabularies start arising for specialized uses?

  - Can NCI keep up with emerging vocabularies or come up with something that leads to that specialized vocabulary?

  - Doesn't necessarily have to be NCI

- All content will need to be reviewed and approved by the VCDE WS

  - VCDE WS needs to identify a minimal set of requirements for semantic rigor

	- We need to allow flexibility in the interim - Burden should fall on the local vocabulary to create extensions to review process - Local dialect, still need common language - Vocabularies in the interim phase may appear and deprecate or get promoted over time ■ Automation of process ■ How do we ensure that low quality, convenient terminology does not become the norm if not approved? ■ At what point in the scientific process does the caBIG become relevant to the provider? ■ We need to establish a low- activation energy from lab notebook to publisher. ■ Local vocabularies need to maintain mapping between local vocabularies and bona fide caBIG approved vocabularies. ■ How do we create a rapid review process so we don't have to worry about inconsistent local terminologies? ■ Infrastructure available at point of terminology entry **caBIG Business Process** This topic was tabled for a future meeting.
Other *(copy & paste from agenda)*	http://cabig.nci.nih.gov/workspaces/VCDE/Meetings/f2f-meetings/index_html/document_view